

Tissue-Wide Expression Profiling Using cDNA Subtraction and Microarrays to Identify Tumor-Specific Genes

Stefan Amatschek,¹ Ulrich Koenig,¹ Herbert Auer,³ Peter Steinlein,⁴ Margit Pacher,⁵ Agnes Gruenfelder,⁵ Gerhard Dekan,² Sonja Vogl,⁵ Ernst Kubista,⁵ Karl-Heinz Heider,⁶ Christian Stratowa,⁶ Martin Schreiber,⁵ and Wolfgang Sommergruber⁶

Departments of ¹Dermatology and ²Pathology, University of Vienna, Vienna, Austria; ³Human Cancer Genetics, Ohio State University, Columbus, Ohio; ⁴Institute of Molecular Pathology, Vienna, Austria; ⁵Department of Obstetrics and Gynecology, University of Vienna and Ludwig Boltzmann Institute of Clinical and Experimental Oncology, Vienna, Austria; and ⁶Boehringer Ingelheim Austria, Vienna, Austria

ABSTRACT

With the objective of discovering novel putative intervention sites for anticancer therapy, we compared transcriptional profiles of breast cancer, lung squamous cell cancer (LSCC), lung adenocarcinoma (LAC), and renal cell cancer (RCC). Each of these tumor types still needs improvement in medical treatment. Our intention was to search for genes not only highly expressed in the majority of patient samples but which also exhibit very low or even absence of expression in a comprehensive panel of 16 critical (vital) normal tissues. To achieve this goal, we combined two powerful technologies, PCR-based cDNA subtraction and cDNA microarrays. Seven subtractive libraries consisting of ~9250 clones were established and enriched for tumor-specific transcripts. These clones, together with ~1750 additional tumor-relevant genes, were used for cDNA microarray preparation. Hybridizations were performed using a pool of 16 critical normal tissues as a reference in all experiments. In total, we analyzed 20 samples of breast cancer, 11 of LSCC, 11 of LAC, and 8 of RCC. To select for genes with low or even no expression in normal tissues, expression profiles of 22 different normal tissues were additionally analyzed. Importantly, this tissue-wide expression profiling allowed us to eliminate genes, which exhibit also high expression in normal tissues. Similarly, expression signatures of genes, which are derived from infiltrating cells of the immune system, were eliminated as well. Cluster analysis resulted in the identification of 527 expressed sequence tags specifically up-regulated in these tumors. Gene-wise hierarchical clustering of these clones clearly separated the different tumor types with RCC exhibiting the most homogenous and LAC the most diverse expression profile. In addition to already known tumor-associated genes, the majority of identified genes have not yet been brought into context with tumorigenesis such as genes involved in bone matrix mineralization (*OSN*, *OPN*, and *OSF-2*) in lung, breast, and kidney cancer or genes controlling Ca^{2+} homeostasis (*RCN1*, *CALCA*, *S100* protein family). *EGLN3*, which recently has been shown to be involved in regulation of hypoxia-inducible factor, was found to be highly up-regulated in all RCCs and in half of the LSCCs analyzed. Furthermore, 42 genes, the expression level of which correlated with the overall survival of breast cancer patients, were identified. The gene dendrogram clearly separates two groups of genes, those up-regulated such as *cyclin B1*, *TGF- β 3*, *B-Myb*, *Erg2*, *VCAM-1*, and *CD44* and those down-regulated such as *MIG-6*, *Esp15*, and *CAK* in patients with short survival time.

INTRODUCTION

Lung, breast, and colorectal cancer are the most common cancers in the industrial world. In 2000, >3 million people were diagnosed to suffer from one of these tumors (1–3). Although improvements in diagnostics and therapy have reduced cancer mortality, still 65% of all

cancer patients die. Prognosis for people suffering from lung cancer is particularly poor, with a mortality rate close to 90%. Two main reasons are responsible for the fact that cancer is expected to become the leading cause of death within a few years: first, cancer is a disease of multiple accumulating mutations that are becoming manifest in human populations with an increasingly prolonged life span (4), and second, neoplastic diseases still have many unmet needs, including lack of understanding of the mechanisms underlying cancer-related deaths and the difficulty in identification of the corresponding risk factors and development of specific targeted molecular therapies. The reason thus lies within the enormous complexity of tumor formation and tumor progression on the molecular level. Many efforts to detect differential gene expression between tumor and normal tissues have discovered numerous differentially expressed genes. However, the mechanistic contribution to tumorigenesis of many of these genes is still unknown.

Because of the genetic instability—characteristic of almost all cancer types—patients suffering from superficially identical tumors show an enormous variability in their gene expression profiles (patient-specific transcription profile). To aim at a more efficient and individual therapy, it will be necessary to distinguish between patient-specific transcription profiles and the altered expression pattern underlying all tumors of the same type. Thus far, the majority of studies applying cDNA microarray technology for the detection of differences in gene expression in cancers have used randomly selected expressed sequence tags and probes derived from a specific tumor and its corresponding normal tissue (5–10). However, these approaches are detrimental to two essential aspects: first, a random selection of cDNAs never covers the entire subset of genes that are specifically overexpressed in cancer, *e.g.*, genes such as differentiation antigens are very unlikely to appear in a random selection of expressed sequence tags. Second, because of missing information about expression profiles of genes in critical (vital) normal tissues, predictions whether a certain gene may serve as a potential clinical target in chemo- and/or immunotherapy are not feasible.

Because of the lack of effective therapies for many major tumor types, the medical need for improved and new approaches of cancer treatment is obvious. Searching for novel targets for tumor therapy is therefore a major goal in this field. In recent publications, the use of cDNA microarrays has been shown to be a powerful tool to detect gene expression differences in cancers (8, 9, 11–13). The cDNA array technique has been also successfully combined with subtractive hybridization to detect tumor-specific transcripts in, for example, lung squamous cell cancer (LSCC) (14). This approach takes advantage of a preselected set of clones that might be more representative than random expressed sequence tags.

The present work is aimed at the identification of novel putative intervention sites for anticancer therapy using a combination of subtractive hybridization and cDNA microarray technology. Importantly, we included a set of 22 samples from various normal tissues to allow discrimination between tumor-specific genes and those that are also

Received 7/17/03; revised 11/16/03; accepted 11/26/03.

Grant support: GEN-AU Program by funds of the Austrian Ministry of Education, Science, and Arts.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Requests for reprints: Wolfgang Sommergruber, Boehringer Ingelheim Austria, Department of NCE Lead Discovery, Dr. Boehringer-Gasse 5-11, Vienna, A-1120, Austria. Phone: 43-80105-2399; Fax: 43-80105-2782; E-mail: wolfgang.sommergruber@vie.boehringer-ingelheim.com.

expressed in vital normal tissues. This approach allowed us to identify genes expressed exclusively in tumors but not in a comprehensive panel of vital normal tissues, which is a prerequisite in the design of a more specific anticancer immuno- or chemotherapy with less side effects. In our experimental approach, we focused on the transcriptional profiling of lung (squamous and adeno), breast, and renal cell cancer (RCC). Seven cDNA libraries were generated by subtracting cDNA fragments derived from normal tissues or primary cell lines from corresponding tumor tissues or tumor cell lines. Subsequently, the derived tumor-enriched clone collection (~9250 clones in total), together with ~1750 additional tumor-relevant genes, was used for the production of cDNA arrays. In extensive hybridization experiments, 50 different tumor samples were analyzed and compared with the expression profile of 22 different normal tissues. This approach allowed the selection of genes that show a significantly higher expression level in tumor tissues than in any of the analyzed normal tissues. To our knowledge, this is the first article of a tissue-wide expression profile used to increase the significance of expression data of tumor samples. Furthermore, a subset of differentially regulated genes was identified that can be correlated with poor prognosis in breast cancer patients.

MATERIALS AND METHODS

RNA Preparation and Quality Control. All tumor tissue samples were collected at the University of Vienna and the University of Graz following approval by their Institutional Review Boards and written informed consent. Tissues were snap frozen in liquid nitrogen immediately after surgical resection and stored at -80°C . For isolation of total RNA, serial cryosections were directly dissolved in 4 M guanidinium thiocyanate containing 1% β -mercaptoethanol, and the lysate was subjected to ultracentrifugation over a CsCl gradient (15). The first and the last cryosections (6 μm) of the biopsies were used for standard H&E staining to verify integrity and tumor cell content of analyzed tissues. Poly(A)⁺ RNA (mRNA) was extracted using the Oligotex Direct mRNA kit (Qiagen, Hilden, Germany) according to the manufacturer's instructions. mRNAs of normal tissues were purchased from Clontech (Palo Alto, CA), Invitrogen (Carlsbad, CA), and Ambion (Austin, TX). All RNAs were analyzed with the Agilent 2100 Bioanalyzer and the RNA 6000 Nano Assay kit (Agilent Technologies, Palo Alto, CA) to determine RNA quality and quantity according to the manufacturer's protocol. To test cDNA samples derived from tumor material for residual genomic DNA, PCR was performed with 0.1 μg of cDNA and primers specific for an intronic region of glyceraldehyde-3-phosphate dehydrogenase (*GAPDH*): *GAPDH* intron forward: 5'-CGCGTCTACGACCTTGCGGGCT-3'; and *GAPDH* intron reverse: 5'-GCTTTCCTAACGGCTGCCCATCA-3'. The integrity of synthesized cDNA was determined by PCR using *GAPDH* exon-specific primers (*GAPDH* exon forward: 5'-AAGGTGAAGGTCGGAGTCAACG-3'; and *GAPDH* exon reverse: 5'-GGCAGAGATGATGACCCTTTGGC-3'). An aliquot of normal and tumor tissue poly(A)⁺ RNA was transcribed into Cla-oligo(dT)-primed (5'-ATTCGCGACTGATGATCGAT₍₁₆₎-3') cDNA. By linear amplification from the 3'-end of cDNAs with a Cla-specific primer (5'-ATTCGCGACT-

GATGATCGAT-3'), the samples were analyzed for degradation of mRNA. Intact mRNA results in a heterogeneous distribution from ~8 to 2 kb on agarose gels.

Construction of Subtractive Libraries. All libraries were generated by suppression subtraction hybridization (SSH), using poly(A)⁺ RNA of tumor tissue or a tumor cell line as a tester, and the corresponding normal tissue or a pool of normal tissues or a primary cell line as a driver. A modified protocol of the PCR-Select cDNA Subtraction kit was used (Clontech). As a major difference to the supplier's protocol, cDNA was not digested with restriction enzyme *RsaI* but with a pool of six restriction enzymes (5 units each of *EcoRV*, *NaeI*, *NruI*, *ScaI*, *SspI*, and *StuI*) to increase average length of cDNAs; cDNA was first incubated 1.5 h in buffer A (Promega, Madison, WI) with *NaeI* and *StuI*. After increase of the NaCl concentration from 6 to 150 mM, cDNA was incubated for another 1.5 h at 37°C with *EcoRV*, *NruI*, *ScaI*, and *SspI*. All subsequent steps were performed according to the protocol of the PCR-Select cDNA Subtraction kit (Clontech). After DNA amplification, PCR products were cloned into pCR 2.1-TA vector (Invitrogen).

Preparation of cDNA Microarrays. In addition to clones derived from subtracted libraries, also individually chosen clones with a known or suspected role in tumor formation were used for our microarrays. Therefore, specific primers were synthesized (genXpress) for 105 expressed sequence tags found to be up-regulated in colon cancer (N. Schweifer, personal communication). The corresponding cDNA fragments were amplified from a mixture of placenta and testis cDNA as a template. PCR products corresponding to partial cDNA fragments were cloned into pGEM-T (Promega) and sequenced. Additional 1682 cancer-relevant clones were obtained from the I.M.A.G.E. consortium. In total, 11,040 clones were used for preparation of microarrays (Table 1). All cloned cDNA fragments were amplified by PCR, purified via ethanol precipitation in 96-well plates, and analyzed on agarose gels (16). PCR products were spotted on poly-L-lysine-coated glass slides (Menzel, Braunschweig, Germany) by a customized robotic arrayer (Promedia Associates, New York, NY). Spotted DNA was cross-linked to the glass surface of the chips by UV irradiation (60 mJ). Chips were then blocked in blocking solution (1.37 g of succinic anhydride in 10 ml of 0.2 M sodium borate and 90 ml of *N*-methyl-pyrrolidone), and double strands were denatured in boiling water.

Labeling and Hybridization. Hybridization was performed as previously described by Eisen and Brown (16). As a reference probe in the two-color hybridizations performed, an equal mixture of 16 critical (vital) normal tissues (rectum, bone marrow, lymph node, spleen, skeletal muscle, small intestine, thymus, trachea, brain, heart, kidney, liver, lung, pancreas, spleen, stomach, and colon) was used. Briefly, 500 ng of poly(A)⁺ RNA were subjected to oligo(dT)-primed reverse transcription following the instructions of the Superscript Reverse Transcription kit (Clontech). The reaction was carried out in a final volume of 40 μl at 42°C. Fluorescent nucleotides Cy5-dUTP and Cy3-dUTP (Amersham Pharmacia, Piscataway, NJ) were used at 0.1 mM. The nucleotide concentrations were 0.5 mM dGTP, dATP, dCTP and 0.2 mM dTTP. Two μl of Superscript II (200 units/ μl ; Invitrogen) were added at the beginning of the labeling reaction, and 1 additional μl was added after 1 h and the reaction continued for an additional hour. Unlabeled RNA was digested using the RNase One kit (Promega) according to the manufacturer's specifications. The Cy5 and Cy3 probes were pooled and, following addition of 15 μg of human Cot1 DNA (Clontech), 3 μg of poly(dA)₄₀₋₆₀, and 6 μg of tRNA (Sigma, St. Louis, MO), subjected to ethanol/ammonium acetate precipitation

Table 1 Tester and driver for generating subtractive cDNA libraries and origin of the 11,040 clones used for microarray production

Library	Tester	Driver	Number	Diversity
A	Calu-6 (lung carcinoma cell line)	WI-38 (fibroblast-like cell from normal embryonic lung)	1140	65.8%
B	A549 (lung carcinoma cell line)	Pooled normal lung tissues	51	88.4%
C	5 pooled lung squamous cell cancer patient samples	11 pooled normal tissues ^a	1931	32.0% ^a
D	4 pooled lung adenocarcinoma patient samples	11 pooled normal tissues ^a	236	64.1% ^a
R	MZ.PC2 (pancreatic carcinoma cell line)	11 pooled normal tissues ^a	3465	64.0% ^a
K	1 renal cell cancer patient sample	2 pooled normal kidneys	1921	38.2%
AKH	1 reduction mammoplasty patient sample	3 pooled normal tissues ^b	509	53.3% ^b
IMAGE	Sequence-verified clones from I.M.A.G.E. consortium ^c		1682	100.0% ^c
Colon	Expressed sequence tags at least 6-fold up-regulated in colon cancer versus normal colon ^d		105	100.0% ^d

^a Bone marrow, heart, kidney, liver, lung, pancreas, skeletal muscle, small intestine, spleen, stomach, and thymus.

^b Placenta, peripheral blood lymphocytes, and skeletal muscle.

^c Most of the selected I.M.A.G.E. clones are known to be relevant for tumor formation.

^d Expressed sequence tags are derived from a study with colon cancer patient samples carried out on Affymetrix GeneChips (N. Schweifer, personal communication).

(15). The pellet was resuspended in 10.5 μ l of water. Six μ l of 20 \times saline-sodium phosphate-EDTA, 1.5 μ l of 50 \times Denhardt's solution, 10.5 μ l of formamide, and 0.75 μ l of 20% SDS were added, and the probe was preincubated for 1 h at 50°C. Chips were prehybridized with 30 μ l of prehybridization buffer [210 μ l of formamide, 120 μ l of 20 \times saline-sodium phosphate-EDTA, 60 μ l of 50 \times Denhardt's solution, 6 μ l of salmon sperm DNA (10

Table 2 Description of normal tissues and tumor samples including TNM grading used for transcription profiling

A. Tumor Tissues				
Sample	Grade and Stage	Histology	cDNA/GAPDH/DNA	Comment
Lung adenocarcinomas (ACs)				
LA 01	Unknown	10% tc	+/+/-	AC
LA 02	T4 G1 N0	50% tc	+/+/-	AC
LA 03	T2 G2 N0	50% tc, > il	+/+/-	AC
LA 04	T2 G2 N1	60% tc	+/+/-	AC
LA 05	T4 G2 N2	70% tc	+/+/-	AC
LA 06	T2 G3 N0	50% tc	+/+/-	Necrotic, atypical
LA 07	T2 G3 N1	80% tc	+/+/-	Necrotic, atypical
LA 08	T4 G3 N1	70% tc	+/+/-	AC
LA 09	T1 G2 N0	50% tc	+/-/-	AC
LA 10	T2 G2 N0	70% tc	+/+/-	AC
LA 11	T2 G2 N0	40% tc	+/+/-	AC
Lung squamous cell carcinoma (SCC)				
LS 01	Unknown	90% tc	+/+/-	SSC
LS 02	T2 G2 N0	70% tc	+/+/-	SSC
LS 03	T2 G2 N0	80% tc, > il	+/+/-	SSC
LS 04	T2 G2 N0	40% tc, > il	+/+/-	SSC
LS 05	T2 G2 N0	80% tc, <5% nc	+/+/-	SSC
LS 06	T2 G3 N0	70% tc	+/+/-	SSC
LS 07	T2 G3 N0	80% tc	+/+/-	SSC
LS 08	T2 G3 N1	60% tc	+/+/-	SSC
LS 09	T4 G3 N2	90% tc	+/+/-	SSC
LS 10	T2 G3 N1	90% tc	+/+/-	AC/SCC
LS 11	T2 G2 N2	50% tc	+/+/-	AC/SCC
Breast invasive ductal carcinoma (IDC)/single probes				
B 17	T1c G2 N1	40% tc, 5% nc	+/+/-	IDC
B 18	T1c G3 N0	30% tc, > il	+/+/-	IDC
B 19	T1c G2 N0	50% tc	+/+/-	IDC
B 20	T2 G2 N1	25% tc, 10% hc	+/+/-	IDC
Renal cell carcinoma (RCC)				
R 01	T1 G1 Nx	90% tc	+/+/-	Clear cell type
R 02	T2 G1 Nx	70% tc	+/+/-	Clear cell type
R 03	T3a G1	90% tc	~/+/-	Clear cell type
R 04	T1 G2 Nx	40% tc	+/+/-	Clear cell type
R 05	T1 G2 Nx	80% tc	+/+/-	Clear cell type
R 06	T3b G3 N0	80% tc	+/+/-	Clear cell type
R 07	T3b G3	90% tc	+/+/-	Clear cell type
R 08	T3a G4	70% tc	+/+/-	Clear cell type
B. Breast carcinoma samples/pairs of primary tumor (AT) and corresponding lymph node metastases (AL)				
Sample	Grade and stage	Histology	Overall survival	Estrogen receptor/ Progesterone receptor
B01 (AT-R85)	T2 G3 N1	IDC	>10 years	-/-
B02 (AL-R86)				
B03 (AT-R87)	T3 G3 N1	IDC	>10 years	-/-
B04 (AL-R88)				
B05 (AT-R89)	T2 G2 N1	IDC	3 years	-/-
B06 (AL-R90)				
B07 (AT-R91)	T2 G2-3 N1	Infiltrating Lobular	2 years	+/-
B08 (AL-R92)				
B09 (AT-R93)	T2 G2 N1	Infiltrating Lobular	>10 years	+/+
B10 (AL-R94)				
B11 (AT-R95)	T2 G3 N1	IDC	7 years	+/+
B12 (AL-R96)				
B13 (AT-R97)	T3 G3 N1	IDC	2 years	-/-
B14 (AL-R98)				
B15 (AT-R99)	T2 G3 N1	IDC	>9 years	+/+
B16 (AL-R100)				

Table 2 Continued

C. Normal tissues				
Sample	Provider	Origin	Age (yrs)	Comments
N01 rectum	InVitrogen	n.a. ^a	n.a.	n.a.
N02 bone marrow	Clontech	Pooled from 74 male/female	18-72	Sudden death
N03 lymph node	Clontech	Pooled from 12 male/female	20-59	Sudden death
N04 skeletal muscle	Clontech	Pooled from 11 male/female	25-59	Sudden death
N05 small intestine	Clontech	Pooled from 12 male/female	29-47	Sudden death
N06 thymus	Clontech	Pooled from 9 female	15-25	Sudden death
N07 trachea	Clontech	Pooled from 84 male/female	17-70	Trauma
N08 brain	Clontech	1 female	15	Trauma
N09 heart	Clontech	Pooled from 10 male/female	21-51	Trauma
N10 kidney	Clontech	Pooled from 6 male/female	28-52	Sudden death
N11 liver	Clontech	1 male	35	Sudden death
N12 lung	Clontech	Pooled from 5 male	25-40	Trauma
N13 pancreas	Clontech	Pooled from 9 male/female	19-75	Sudden death
N14 spleen	Clontech	Pooled from 6 male/female	30-58	Sudden death
N15 stomach	Clontech	1 male	57	Sudden death
N16 colon	Ambion	1 female	34	IC Bleed
N17 uterus	Clontech	n.a.	n.a.	n.a.
N18 mammary gland	Clontech	Pooled from 9 donors	n.a.	Reduction mammoplasty
N19 prostate	Clontech	Pooled from 47 donors	14-57	Sudden death
N20 fetal brain	InVitrogen	Pooled from 5 male/female fetuses	n.a.	Spontaneously aborted
N21 fetal lung	Clontech	Pooled from 9 male/female fetuses	20-25 weeks	Spontaneously aborted
N22 placenta	Clontech	Pooled from 7 donors	20-35	Normal parturition

^a n.a., not available.

For tumor samples the TNM-grade and stage, the percentage of tumor cells [tc], residual normal tissue cells [nc], hyperplastic cells [hc], infiltrating leukocytes (more than 5% of all cells [$>il$]), and comments on histology are stated. Tumor staging was done according to the UICC (53). Results from PCR-analysis of amplification from the 3'-end with Cla-specific primers [cDNA], GAPDH-specific primers [GAPDH], and GAPDH-intron primers [DNA] are classified into negative [-], weak [-], and positive [+] signals. Samples showing RNA degradation (only short transcripts in Cla-PCR), no amplification of GAPDH, or detectable genomic DNA, respectively, were not used for microarray hybridizations. ER/PR refers to the estrogen-(ER) and progesterone-(PR) receptor status of the samples. B01 to B16, odd numbers are primary tumors (AT) and even numbers are lymph node metastases (AL). In each case the poly-A⁺RNA of normal tissues was isolated from the whole corresponding organ. The tissues were derived throughout from Caucasian donors.

μ g/ μ l), 190 μ l of H₂O, and 15 μ l of 20% SDS] for 1 h at 50°C. For hybridization, the probe was added to the array, covered with a coverslip, and placed in a sealed chamber to prevent evaporation. After hybridization at 50°C over night, chips were washed in three solutions of decreasing ionic strength (17).

Scanning, Image Analysis, Data Filtering, and Normalization. Microarrays were scanned with an GenePix 4000A scanner (Axon Instruments, Inc., Union City, CA) at 10- μ m resolution. The signal was converted into 16 bits/pixel resolution, yielding a 65,536 count dynamic range. Image analysis and calculation of feature pixel intensities adjusted for local channel specific background was performed using the GenePix Pro 3.0 software (Axon Instruments, Inc.). With this software, we performed gridding, automated spot detection, manual and automated flagging, as well as background subtraction and normalization. Background-subtracted element signals were used to calculate Cy3/Cy5 ratios. Spots were excluded from additional analysis if the ratio of foreground versus background signal was <2. Each microarray was normalized by scaling according to the GenePix normalization factor such that the median of ratios value is 1. For additional evaluation and statistical analysis, output files were exported to a relational Microsoft Access database.

Statistical Analysis. All statistical analyses were performed with the software package S-Plus (Insightful, Inc.) and R.⁷ Student's *t* test was used to select genes most differentially expressed between each two groups, with significance set at $P < 0.02$. Furthermore, P-chance analysis was used to eliminate false positives. Hereby, a *t*-statistic is assigned to each gene, and *P*s (both unadjusted and adjusted according to Westfall and Young step-down

⁷ Internet address: <http://www.r-project.org>.

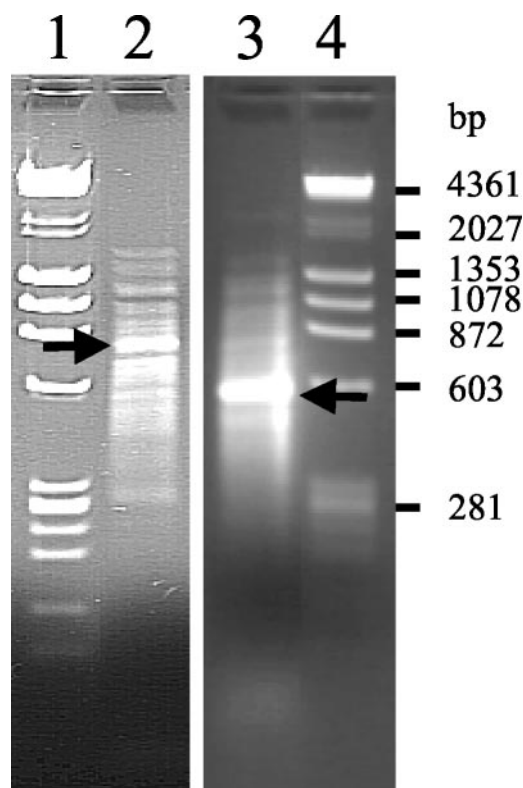


Fig. 1. PCR products of subtractive lung squamous cell carcinoma cDNA libraries generated either by a 4-base or a pool of 6-base recognizing restriction enzymes. Lanes 1 and 4: DNA size marker (Φ \times /HaeIII + λ /HindIII); Lane 2: subtractive library generated using a pool of 6-cutters; Lane 3: subtractive library generated using the 4-cutter *Rsa*I; arrows indicate the respective position for keratin 6A cDNA fragment.

algorithm) are obtained by permutating the samples in place of assuming t-distributions, as described by Dudoit *et al.* (18). To reduce the number of false positives (type I error), only genes with $P < P$ -chance were selected. In addition, only genes with an absolute fold change > 1.5 -fold were chosen for additional analyses. For cluster analysis, either GeneSpring (Silicon Genetics, Redwood City, CA) or Spotfire DecisionSite (Spotfire, Goeteborg, Sweden) were used as software packages.

Quantitative Real-Time PCR. Quantitative real-time PCR was performed in the presence of SYBR Green using the Lightcycler-DNA Master SYBR Green I kit from Roche (Mannheim, Germany). Comparison with housekeeping genes allows relative quantification of monitored genes in different cDNA samples. Briefly, 100 ng of mRNA were converted to cDNA in a total volume of 50 μ l using the Superscript Reverse Transcription kit (Clontech). One μ l of this mixture was used as template for PCR amplification. Thirty-five PCR cycles were performed as follows: 30 s denaturation at 94°C; 30 s annealing at 65°C; and 45 s for elongation at 72°C. All reverse transcription (RT)-PCR reactions were performed on an ABI PRISM 7700 Detector (Perkin-Elmer/Applied Biosystems, Foster City, CA). All plates contained 60 different cDNAs, a dilution series of a plasmid for the gene of interest, and nontemplate controls. Gene-specific primers were used to amplify fragments of ~ 130 bp. All plates were done in duplicates based on which average copy numbers were calculated. Copy numbers were normalized to β -actin using the following primers for amplification: actin-up, 5'-TGTTTCTGCGCAAGTTAGG-3'; and actin-do, 5'-GTCCACCTTCCAGCAGATGT-3'.

RESULTS

Histological Characterization and RNA Extraction. Before RNA preparation, tumor tissues were histologically assessed to ensure homogeneity and integrity of the tumor tissue and to confirm the histological typing of the tumor samples. Cryosections were taken and stained with H&E. The fraction of tumor cells (in most sam-

ples $> 50\%$), residual normal tissue cells, hyperplastic cells, and infiltrating leukocytes were determined (Table 2). Tumor samples with massive leukocyte infiltration and large necrotic areas were excluded from additional analysis (altogether 50% of tumor samples analyzed).

For quantification and quality control, we subjected all RNA preparations to capillary electrophoresis (Bioanalyzer 2100; Agilent Technologies) and PCR analyses for potential chromosomal DNA contamination using intron-specific primers for *GAPDH*. Thirty percent of all RNA preparations were eliminated because of degradation (28S/18S rRNA ratio < 1.3 plus profile analysis of capillary electrophoresis) or contamination with rRNA ($> 10\%$) and/or chromosomal DNA (positive signal of *GAPDH*-intron-specific PCR). Additionally, each of the poly(A)⁺ preparations from individual tumor samples were analyzed for RNA degradation by performing a linear amplification of Cl-oligo(dT)-primed cDNA aliquots (see "Material and Methods"; Table 2). RNA samples meeting those stringent quality criteria are listed in Table 2 and were subsequently used for microarray experiments. Sixteen of the 20 breast cancer samples used represent pairs of primary tumors and corresponding lymph node metastases (Table 2).

Generation of Tumor-cDNA-Enriched Libraries. As a major improvement to conventional SSH protocols, we aimed at the generation of longer cDNA fragments for subtractive hybridization. We used a combination of six different 6-base recognizing restriction enzymes instead of an enzyme recognizing a 4-base motif, as described in the original protocol (14). This led to an average cDNA length of 800 bp instead of a predicted average fragment size of 256 (4^4) bp. The efficiency of adapter ligation and suppression PCR was found to be the same when either a 4-base (*Rsa*I) or our set of 6-base recognizing restriction enzymes were used (data not shown). As shown in Fig. 1, our approach results in a considerable increase in the average length of cDNA fragments. A prominent band in Fig. 1 was identified by gel isolation and sequencing to be keratin 6A. This keratin 6A cDNA fragment was 850 bp long when our combination of 6-base cutters was applied, but only 580 bp when *Rsa*I was used. Furthermore, sequencing of several hundreds of our clones revealed an average length of ~ 800 bp. The enrichment for known tumor markers such as keratin 6A (19) in the subtractive LSCC library provided evidence for successful subtractive hybridization; a high proportion of sequenced clones of this library were found to be members of the cytokeratin protein family (Table 3). For description of libraries and number of clones that have been used for microarray experiments, see Table 1.

Microarray Experiments, Statistics, and Data Confidence. In total, we derived 9253 clones by SSH from seven different sources

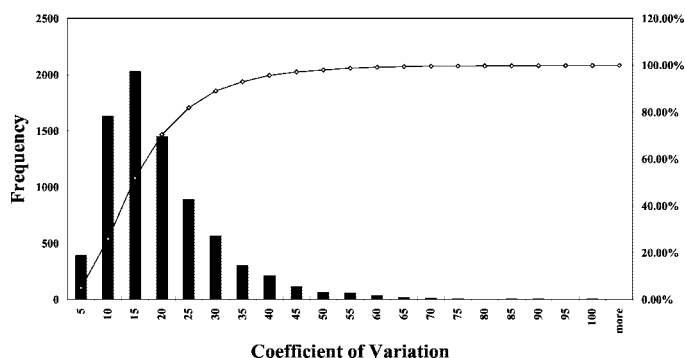


Fig. 2. Histogram of the distribution of coefficients of variation (CV). Data are from a hybridization that was repeated four times under the same conditions. Only spots with valid signals in each of the four hybridizations (71%) were included for calculation. Ninety-five percent of all spots have CVs smaller than 37%, $> 99\%$ of all spots displayed CVs of $< 57\%$.

Table 3 List of genes selectively over-expressed in breast carcinoma, LACs, LSCCs, and RCCs

Tumor type	Gene name	Accession no.	k/n	m	
Breast cancer	Secreted protein, acidic, cysteine-rich (osteonectin) (SPARC)	NM_003118	14/20	8	
	Collagen, type I, α 1 (COL1A1)	NM_000088	13/20	12	
	Collagen, type I, α 2 (COL1A2)	NM_000089	12/20	5	
	Collagen, type III, α 1 (COL3A1)	NM_000090	11/20	9	
	Fibronectin 1 (FN1), transcript variant 1	NM_002026	10/20	8	
	N-acetyltransferase 1 (NAT1)	NM_000662	10/20	1	
	Osteoblast specific factor 2 (fascin I-like) (OSF-2)	NM_006475	9/20	1	
	5T4 oncofetal trophoblast glycoprotein (5T4)	NM_006670	8/20	1	
	Thrombospondin 2 (THBS2)	NM_003247	8/21	1	
	KIAA0225 protein (KIAA0225)	D86978	8/22	1	
	Pre-B-cell leukemia transcription factor 1 (PBX1)	NM_002585	7/20	1	
	Collagen, type VI, α 3 (COL6A3)	NM_057167	7/20	7	
	Platelet-derived growth factor receptor, β polypeptide (PDGFRB)	NM_002609	6/20	1	
	Similar to glucosamine-6-sulfatases (SULF2)	NM_018837	6/20	1	
	Matrix metalloproteinase 11 (stromelysin 3) (MMP11)	NM_005940	6/20	1	
	Interferon, α -inducible protein (GIP3), transcript variant 3	NM_022873	6/20	3	
	Transducin (β)-like 1 (TBL1)	NM_005647	6/20	1	
	Fer-1 (<i>C. elegans</i>)-like 3 (myoferlin) (FER1L3)	NM_013451	6/20	1	
	Matrix metalloproteinase 13 (collagenase 3) (MMP13)	NM_002427	5/20	1	
	Cyclin D1 (PRAD1; parathyroid adenomatosis 1) (CCND1)	NM_053056	5/20	1	
	Melanophilin (MLPH)	NM_024101	5/20	1	
	Nonmetastatic cells 1, protein (NM23A)	NM_000269	5/20	2	
	Chondroitin sulfate proteoglycan 2 (versican) (CSPG2)	NM_004385	5/20	3	
	Prolactin receptor (PRLR)	NM_000949	5/20	1	
	Small inducible cytokine subfamily A (Cys-Cys), member 19 (SCYA19)	NM_006274	5/20	1	
	Homo sapiens H3 histone, family 3B (H3F3B)	NM_005324	4/20	1	
	Stanniocalcin 2 (STC2)	NM_003714	4/20	1	
	Transcription factor AP-2 β (TFAP2B)	NM_003221	4/20	1	
	X-box binding protein 1 (XBP1)	NM_005080	4/20	1	
	Cathepsin K (pyncnodysostosis) (CTSK)	NM_000396	4/20	1	
	Protease, serine, 11 (IGF binding) (PRSS11)	NM_002775	4/20	1	
	Activated RNA polymerase II transcription cofactor 4 (PC4)	NM_006713	4/20	1	
	Chromosome 1 open reading frame 29 (C1orf29)	NM_006820	4/20	1	
	Collagen, type V, α 1 (COL5A1)	NM_000093	4/20	1	
	Hypothetical protein IMPACT (IMPACT)	NM_018439	4/20	1	
	Melanoma differentiation-associated protein-5 (MDA5)	NM_022168	4/20	1	
	Nonmetastatic cells 2, protein (NM23B)	NM_002512	4/20	1	
	Plasminogen activator, urokinase (PLAU)	NM_002658	4/20	1	
	Bone marrow stromal cell antigen 2 (BST2)	NM_004335	4/20	1	
	Lung adenocarcinoma	Ornithine decarboxylase 1 (ODC1)	NM_002539	4/11	13
		Surfactant, pulmonary-associated protein A2 (SFTPA2)	NM_006926	4/11	6
		Transmembrane 4 superfamily member 1 (TM4SF1)	NM_014220	3/11	3
		SHC (Src homology 2 domain-containing) transforming protein 1 (SHC1)	NM_003029	3/11	1
		Solute carrier family 34 (sodium phosphate), member 2, (SLC34A2)	NM_006424	3/11	2
		Chromosome 8 open reading frame 4 (C8orf4)	NM_020130	3/11	3
		Prostaglandin-endoperoxide synthase 2	NM_000963	3/11	1
		Pronapsin A (NAP1)	NM_004851	3/11	1
Dual specificity phosphatase 6 (DUSP6)		NM_022652	3/11	2	
Aspartate beta-hydroxylase (ASPH)		NM_032466	3/11	2	
Chitinase 3-like 1 (cartilage glycoprotein-39), (CHI3L1)		NM_001276	3/11	1	
Plasminogen activator, urokinase (PLAU)		NM_002658	3/11	6	
Reticulocalbin 1, EF-hand calcium binding domain (RCN1)		NM_002901	2/11	1	
Secreted protein, acidic, cysteine-rich (osteonectin) (SPARC)		NM_003118	2/11	2	
Solute carrier family 2 (facilitated glucose, transporter), member 1 (SLC2A1)		NM_006516	2/11	2	
TAF4 RNA polymerase II, TATA box binding protein-associated factor (TAF4)		NM_003185	2/11	1	
Trefoil factor 3 (intestinal) (TFF3)		NM_003226	2/11	2	
Tyrosyl-tRNA synthetase (YARS)		NM_003680	2/11	1	
Tissue factor pathway inhibitor 2 (TFPI2)		NM_006528	2/11	1	
Collagen, type I, α 2 (COL1A2)		NM_000089	2/11	1	
Calcitonin/calcitonin-related polypeptide, α (CALCA)		NM_001741	2/11	1	
Calumenin (CALU)		NM_001219	2/11	1	
Cytochrome P450, subfamily I (CYP1B1)		NM_000104	2/11	1	
Chondroitin sulfate proteoglycan 2 (versican) (CSPG2)		NM_004385	2/11	1	
Collagen, type III, α 1 (COL3A1)		NM_000090	2/11	1	
Cystatin B (stefin B) (CSTB)		NM_000100	2/11	1	
5-hydroxytryptamine (serotonin) receptor 2B (HTR2B)		NM_000867	2/11	1	
Epiregulin (EREG)		NM_001432	2/11	2	
Keratin 6A (KRT6A)		NM_005554	2/11	41	
Lung squamous cell cancer		Keratin 6A (KRT6A)	NM_005554	9/11	80
		RAN, member RAS oncogene family (RAN)	NM_006325	8/11	1
		Neurotrophic tyrosine kinase, receptor, type 1 (NTRK1)	NM_002529	8/11	1
		Solute carrier family 2 (facilitated glucose, transporter), member 1 (SLC2A1)	NM_006516	8/11	6
	Bullous pemphigoid antigen 1 (230/240kD) (BPAG1)	NM_001723	8/11	1	
	S100 calcium-binding protein A2 (S100A2)	NM_005978	8/11	10	
	CDNA FLJ33151 fis	AK057713	7/11	1	
	Aldo-keto reductase family 1, member C3 (AKR1C3)	NM_003739	7/11	3	
	Phosphoglycerate kinase 1 (PGK1)	NM_000291	7/11	1	
	P53-induced protein PIGPC1 (PIGPC1)	NM_022121	7/11	8	
Collagen, type I, α 1 (COL1A1)	NM_000088	7/11	3		

Table 3 *Continued*

Tumor type	Gene name	Accession no.	k/n	m
	<i>N</i> -myc downstream regulated (NDRG1)	NM_006096	7/11	4
	Sperm-specific antigen 2 (SSFA2)	NM_006751	7/11	1
	Desmoplakin (DPI, DPII) (DSP)	NM_004415	7/11	1
	Tumor protein p73-like (TP73L)	NM_003722	6/11	2
	KIAA2019 protein (KIAA2019)	AB095939	6/11	1
	Keratin 6B (KRT6B)	NM_005555	6/11	3
	Tripartite motif-containing 29 (TRIM29)	NM_012101	6/11	1
	Aldo-keto reductase family 1, member B10 (aldose reductase) (AKR1B10)	NM_020299	6/11	1
	Glycoprotein (transmembrane) nmb (GPNMB)	NM_002510	5/11	3
	Keratin 5 (KRT5)	NM_000424	5/11	2
	Keratin 14 (KRT14)	NM_000526	5/11	2
	Osteoblast specific factor 2 (fascin I-like) (OSF-2)	NM_006475	5/11	1
	Hypothetical protein MGC5306 (MGC5306)	NM_024116	5/11	1
	Plasminogen activator, urokinase (PLAU)	NM_002658	5/11	6
	Translational activator GCN1	U88836	5/11	1
	S100 calcium-binding protein A11 (calcizzarin) (S100A11)	NM_005620	5/11	2
	Parathyroid hormone-like hormone (PTH LH)	NM_002820	5/11	2
	Secreted protein, acidic, cysteine-rich (osteonectin) (SPARC)	NM_003118	5/11	3
	Disintegrin and metalloproteinase domain 9 (meltrin γ) (ADAM9)	NM_003816	5/11	1
	Aldo-keto reductase family 1, member C1 (AKR1C1)	NM_001353	5/11	2
	Aldo-keto reductase family 1, member C2 (AKR1C2)	NM_001354	5/11	3
	Transmembrane protein vezatin (VEZATIN)	NM_017599	5/11	1
	Chromosome 18, clone RP11-650P15	AC021549	5/11	3
	Solute carrier family 5 (sodium-dependent vitamin transporter), member 6 (SLC5A6)	NM_021095	4/11	1
	Peroxisome oxidin 1 (PRDX1)	NM_002574	4/11	1
	Secreted phosphoprotein 1 (osteopontin) (SPP1)	NM_000582	4/11	2
	Aspartate β -hydroxylase (ASPH)	NM_032466	4/11	2
	Laminin, γ 2 (LAMC2)	NM_005562	4/11	2
	Claudin 1 (CLDN1)	NM_021101	4/11	1
	Annexin A1 (ANXA1)	NM_000700	4/11	1
	Collagen, type I, α 2 (COL1A2)	NM_000089	4/11	1
	Collagen, type III, α 1 (COL3A1)	NM_000090	4/11	6
	Cystatin B (stefin B) (CSTB)	NM_000100	4/11	1
	Desmoglein 3 (pemphigus vulgaris antigen) (DSG3)	NM_001944	4/11	1
	Dual specificity phosphatase 5 (DUSP5)	NM_004419	4/11	1
	EGL nine (<i>C. elegans</i>) homologue 3 (EGLN3)	NM_022073	4/11	1
	Glyoxalase I (GLO1)	NM_006708	4/11	1
Renal cell cancer	EGL nine (<i>C. elegans</i>) homologue 3 (EGLN3)	NM_022073	8/8	1
	<i>N</i> -myc downstream regulated (NDRG1)	NM_006096	8/8	22
	Vascular endothelial growth factor (VEGF)	NM_003376	8/8	20
	Insulin-like growth factor binding protein 3 (IGFBP3)	NM_000598	8/8	32
	Endothelial cell-specific molecule 1 (ESM1)	NM_007036	8/8	1
	Met proto-oncogene (hepatocyte growth factor receptor) (MET)	NM_000245	7/8	7
	Human DNA sequence from clone RP11-269F20 on chromosome 1	AL591721	7/8	1
	Regulator of G-protein signalling 5 (RGS5)	NM_003617	6/8	16
	Alpha glucosidase II α subunit (G2AN)	NM_014610	6/8	1
	Interleukin enhancer binding factor 2, 45kD (ILF2)	NM_004515	6/8	1
	Epidermal growth factor receptor (EGFR)	NM_005228	6/8	2
	Transcription factor 19 (SC1) (TCF19)	NM_007109	6/8	1
	Yes-associated protein 1, 65 kDa (YAP1)	NM_006106	6/8	1
	Phosphodiesterase IB, calmodulin-dependent (PDE1B)	NM_000924	5/8	1
	MSTP032 protein (MSTP032)	NM_025226	5/8	11
	Aldehyde dehydrogenase 1 family, member A3 (ALDH1A3)	NM_000693	5/8	1
	Secreted protein, acidic, cysteine-rich (osteonectin) (SPARC)	NM_003118	4/8	6
	Platelet-derived growth factor receptor, β polypeptide (PDGFRB)	NM_002609	4/8	1
	Interferon-induced protein with tetratricopeptide repeats 1 (IFIT1)	NM_001548	4/8	1
	Human DNA sequence from clone 596C15 on chromosome Xq23	AL031387	4/8	1
	Cyclin D1 (PRAD1; parathyroid adenomatosis 1) (CCND1)	NM_053056	4/8	1
	Plasminogen activator, urokinase (PLAU)	NM_002658	3/8	1
	Phosphoglycerate kinase 1 (PGK1)	NM_000291	3/8	4
	Inhibin, beta B (activin AB β polypeptide) (INHBB)	NM_002193	3/8	2
	Hypothetical protein FLJ13081 (FLJ13081)	NM_024834	3/8	1
	Heat shock protein 75 (TRAP1)	NM_016292	3/8	1
	BAC clone RP11-529E15	AC073218	3/8	1
	Chromosome 1 open reading frame 8 (C1orf8)	NM_004872	3/8	1
	Transforming growth factor, α (TGFA)	NM_003236	2/8	1
	Hypothetical protein DKFZp434F0318 (DKFZP434F0318)	NM_030817	2/8	1
	Transmembrane 4 superfamily member 1 (TM4SF1)	NM_014220	2/8	1
	CD27-binding (Siva) protein (SIVA)	NM_006427	2/8	1
	Integrin, β 1 (ITGB1)	NM_002211	2/8	1
	Transforming growth factor, β -induced, 68kD (TGFB1)	NM_000358	2/8	4
	Disintegrin and metalloproteinase domain 9 (meltrin γ) (ADAM9)	NM_003816	2/8	2
	Tissue factor pathway inhibitor 2 (TFPI2)	NM_006528	2/8	1
	Cytochrome P450 (CYP2J2)	NM_000775	2/8	1
	Fer-1 (<i>C. elegans</i>)-like 3 (myoferlin) (FER1L3)	NM_013451	2/8	1
	Tumor necrosis factor receptor superfamily, member 21 (TNFRSF21)	NM_014452	2/8	1

All genes listed exhibit at least a two-fold up-regulation compared to the highest expression value in any critical normal tissue in at least 20% of a given tumor type. Additionally, the number of times a gene was detected in the libraries (m) and the frequency by which a gene was at least two-fold up-regulated in a specified tumor type (k/n) is shown.

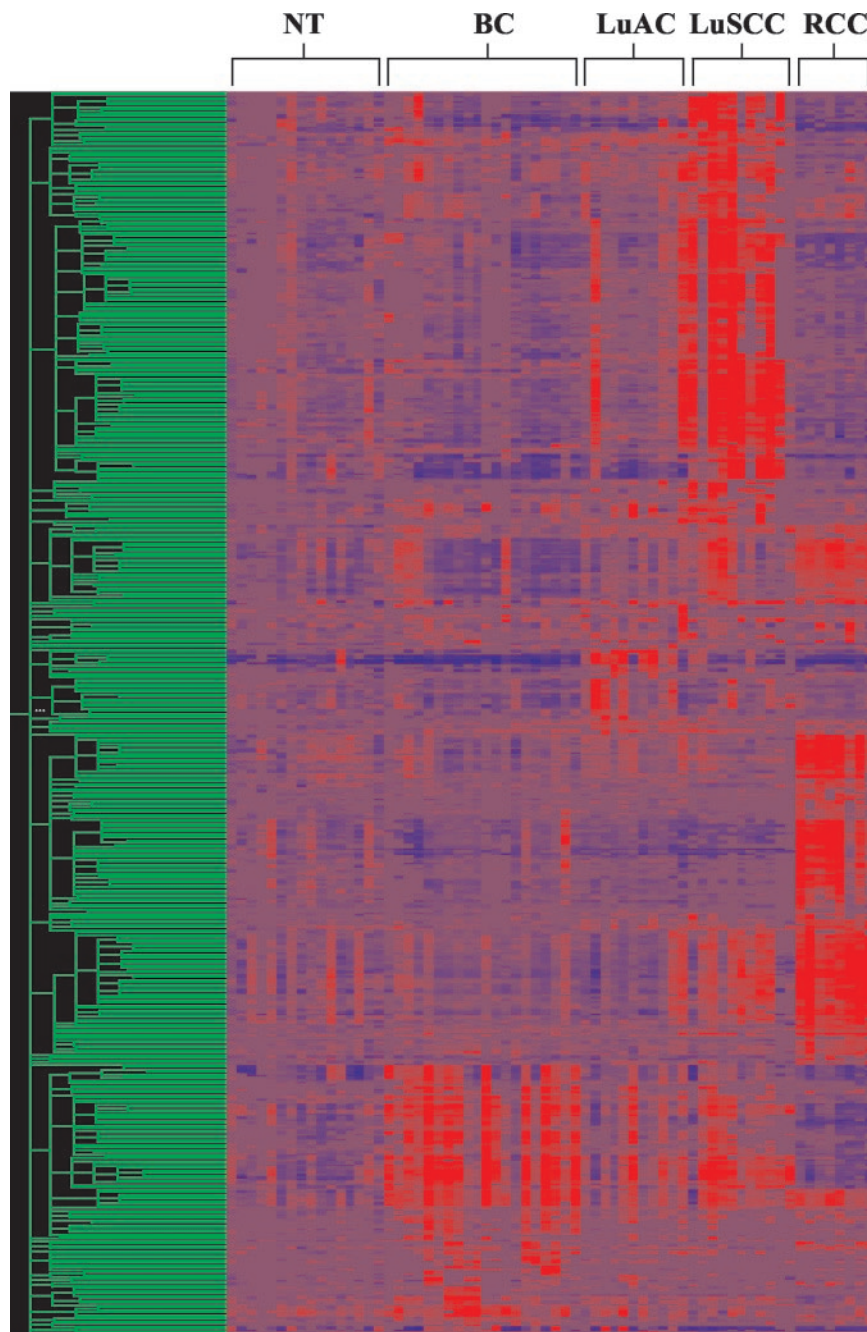


Fig. 3. Hierarchical clustering on a set of 527 clones. NT, vital normal tissues (for list see "Materials and Methods"); BC, breast cancer; LuAC, lung adenocarcinoma; LuSCC, lung squamous cell carcinoma; and RCC, renal cell cancer. *Red blocks* represent up-regulation and *blue ones* down-regulation. In particular, RCC, LuSCC and breast cancers form well-defined clusters.

and 105 clones by RT-PCR cloning of individual genes, which had been shown previously to be at least 6-fold up-regulated in colon tumors; 1682 sequence-verified and tumor-relevant genes were obtained from the I.M.A.G.E. consortium (Table 1). cDNA microarrays were produced from these 11,040 clones by robotic arraying onto poly-L-lysine-coated glass slides. Fifty selected tumor samples [11 lung adenocarcinomas (LACs), 11 LSCCs, 20 breast carcinomas, and 8 RCCs], 16 critical (vital) normal tissues (rectum, bone marrow, lymph node, skeletal muscle, small intestine, thymus, trachea, brain, heart, kidney, liver, lung, pancreas, spleen, stomach, and colon) and 6 noncritical normal tissues (uterus, breast, prostate, fetal brain, fetal lung, and placenta) were analyzed. As a reference probe in the two-color hybridizations performed, an equal mixture of the 16 critical normal tissues was used. Fluorescence scanning and image analysis resulted in ratios of Cy3/Cy5 intensities that were used for additional analyses.

For comparison of individual hybridizations with each other, data were normalized before data mining. As our chips comprised 11,040 clones of nine different sources (Table 1), we regarded our clone collection to be sufficiently diverse and balanced. Accordingly, we applied a global normalization on our data to balance total intensities of Cy3 (red) and Cy5 (green) fluorescence by a linear transformation. By this procedure, ratios were divided with a ratio of global Cy3/Cy5 intensities of each chip.

We next performed statistical analysis on the confidence and extent of variation of our data (Fig. 2). For that purpose, one hybridization was repeated four times under the same conditions. More than 95% of all spots displayed coefficients of variation (coefficient of variation = SD in percentage of average) smaller than 37%, whereas >99% of all spots displayed coefficient of variations of <57%. Thus, with a probability of 99%, any Cy3/Cy5 ratio determined has a maximum error of 57%. Therefore, 2-fold changes are statistically

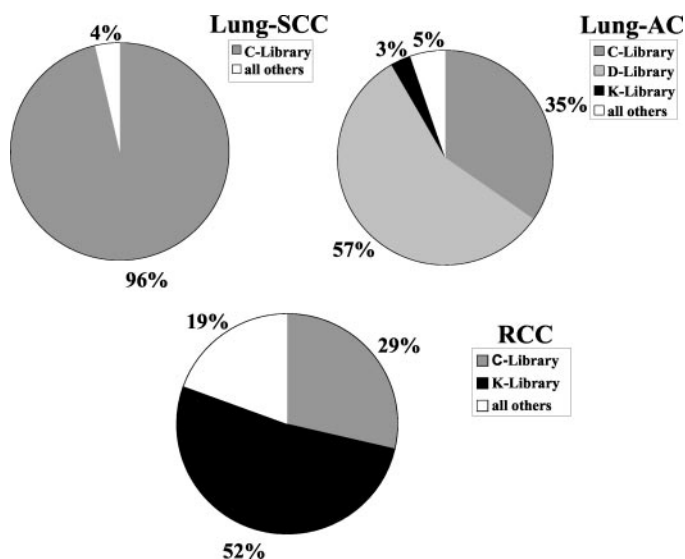


Fig. 4. Distribution of clones among subtractive libraries. The 100 most up-regulated clones for lung-squamous cell carcinoma (SCC), lung-adenocarcinoma (AC), and renal cell cancer (RCC), respectively, were analyzed with reference to their origin from the different subtractive libraries (for description of libraries, see Table 1), e.g., 96% of the most up-regulated clones in lung-SCC were derived from the lung-SCC subtractive library (C-Library).

significant for >99% of all spots. Values with higher intensities are more reliable, indicated by a smaller coefficient of variation (data not shown). Accordingly, spots that did not show intensities at least 2-fold higher than the background were excluded from subsequent analyses.

Cluster Analysis. On the basis of normalized and prefiltered data, we first determined N_{\max} for each gene, i.e., the highest expression value in any of the 16 critical normal tissues. All normalized expression values of the 11,040 genes in 22 normal tissues and 50 tumor samples are listed in Supplemental Data, Table 1. Subsequently, all genes were selected that showed at least 2-fold higher expression in at least 20% of samples of any tumor type compared to N_{\max} . Applying these restrictive criteria resulted in a list of 527 clones representing 130 different genes, 116 of which coded for proteins of known functions (summarized in Table 3).

Gene-wise hierarchical clustering of these 527 clones clearly separated the different tumor types (Fig. 3). This was not necessarily to be expected because these 527 clones were selected based on differential expression between tumors and normal tissues but not on differential expression between different tumor types. However, the extent of similarity of tumors from the same tissue of origin varied strongly (Fig. 3). All 8 RCCs showed nearly identical expression profiles, whereas expression profiles of LACs were remarkably variable among different patients. Sample-wise clustering confirmed these data, as RCCs again clustered together most closely (data not shown).

Additional analysis of the 527 clones (Table 3) revealed that at least half of the clones selected as up-regulated in a given tumor type were derived from the corresponding subtractive library (Fig. 4); for example, 96% of the most highly up-regulated genes in LSCC have been derived from the library based on subtraction of normal tissues from LSCC tissues. A very similar situation is found with the differentially regulated genes in LAC and RCC; >50% of these genes are derived from their corresponding subtractive libraries (Fig. 4). These data demonstrate the specificity of the subtractive libraries and underscore that the combination of subtractive hybridization and cDNA microarray technology is a highly efficient way to identify differences in gene expression

profiles. Furthermore, each clone of a given subtractive library is representative of the tumor tissue where it is derived from.

Quantitative Real-Time PCR. To verify the results of microarray experiments, real-time PCR experiments were performed. Sixty of the total 72 RNA samples were subjected to reverse transcription and subsequent quantification in real-time PCR. β -Actin was used as a reference because it exhibits the most constant expression level in all our tissues analyzed with microarrays.

Fig. 5 shows a side-by-side comparison of expression levels measured by microarray analysis and real-time PCR for EGLN3, a gene that was found to be highly up-regulated in all RCCs. Real-time PCR and microarray analyses show nearly identical expression patterns. No significant expression ("flagged spots") in microarray experiments (absent bars in Fig. 5A) correlate well with absence or low expression levels in real-time PCR (Fig. 5B). Overexpression of EGLN3 in RCCs was also confirmed by immunohistochemistry (20). Four additional genes, *NDGRI*, *OSF-2*, *TP73L*, and *NATI* were verified in the same way, all exhibiting high agreement between microarray and real-time PCR results. Linear regression analysis revealed R^2 values of 0.64 (*EGLN3*), 0.69 (*NDGRI*), 0.61 (*OSF-2*), 0.87 (*TP73L*), and 0.67 (*NATI*).

Thus far, two splice variants have been demonstrated for *EGLN3* (formerly described as 9D7), which recently has been shown to be involved in the regulation of the hypoxia-inducible factor (20, 21). By RT-PCR studies using RNA samples derived from normal tissues and tumor samples, we demonstrated that the shorter splice variant is dominant over the longer one in all tissues. In addition, the shorter splice variant is up-regulated in all RCCs and half of LSCCs (data not shown).

Survival Correlation. Sixteen of our breast cancer samples represent 8 pairs of primary tumors and lymph node metastases. Three patients succumbed to their disease within 3 years after diagnosis, and 4 patients were still alive at least 9 years after diagnosis. This allowed us to classify the patients into two groups and apply a statistical filter criterion to select the genes that are most significantly correlated with overall survival. One patient had an overall survival of 7 years and was not included in this analysis.

Starting with the expression values from the complete set of 11,040 clones, this approach resulted in a list of 45 clones that correlated best with the overall survival, corresponding to 42 genes; 3 genes were represented twice (Fig. 6). We next subjected the expression profile of these 42 genes to two distinct unsupervised cluster methods, hierarchical clustering (Fig. 6), and principal component analysis (data not shown). Both methods were able to correctly separate the samples into two distinct groups. Furthermore, the gene dendrogram clearly separates two groups of genes: those overexpressed in patients with short survival time and those down-regulated in patients with short survival. The former group includes cyclin B1, *TGF- β 3*, the transcription factors *Erg2* and *B-Myb*, and the cell adhesion molecules *VCAM-1* and *CD44*, whereas genes down-regulated in patients with short survival include *MIG-6*, *Eps15*, and *CAK*.

DISCUSSION

Generation of a Large, Extensively Sequenced Clone Collection of Tumor-Specific Genes Optimized for cDNA Microarray Production. We have combined two powerful gene expression profiling technologies, SSH and cDNA microarrays, for analysis of differential gene expression in human tumors. A combination of subtractive cDNA cloning and cDNA microarrays has previously been successfully used for the rapid identification of differentially expressed genes in estrogen receptor-positive and -negative breast cancer cell lines (22) in human prostate cancer (23), LSCC (24),

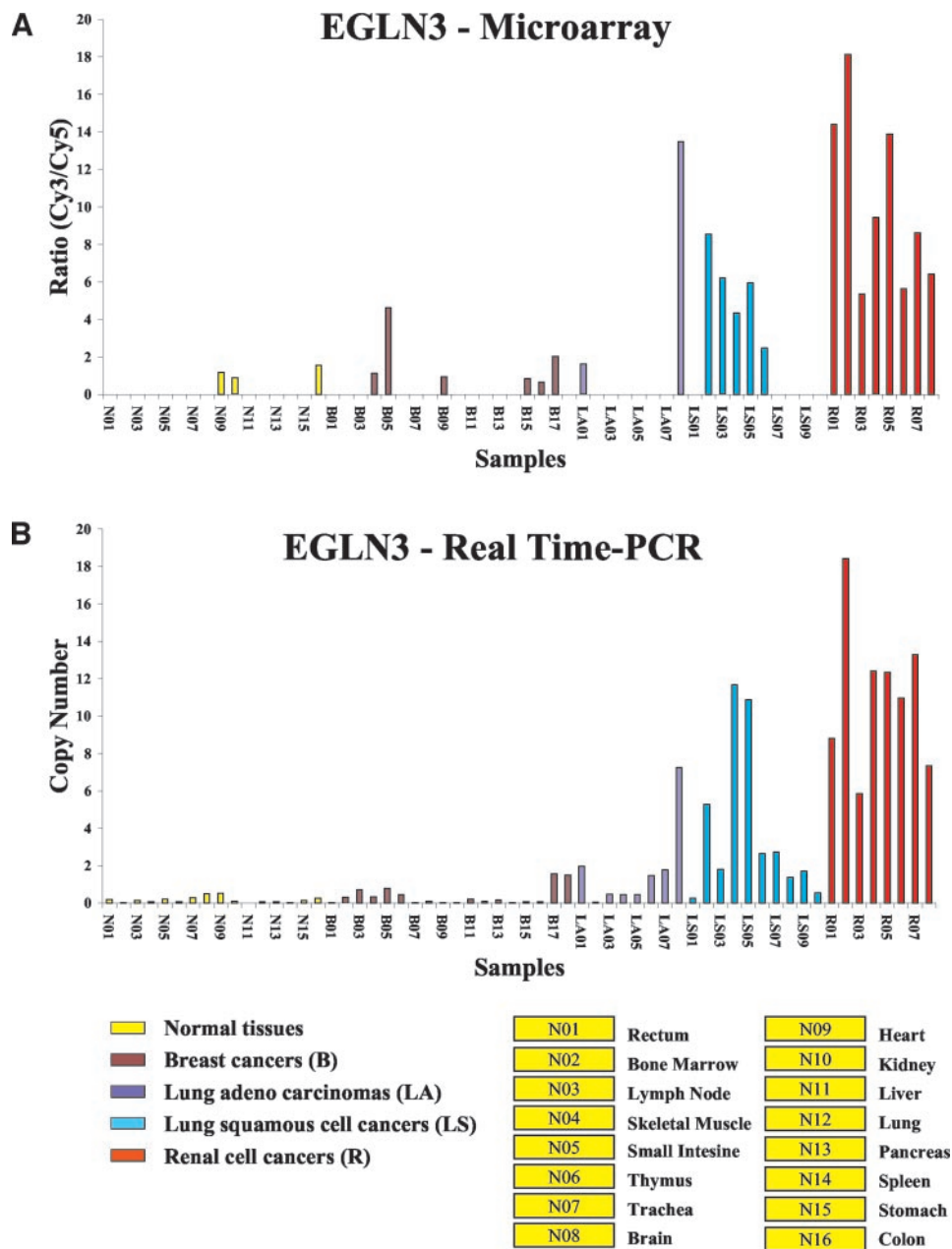


Fig. 5. Comparison of expression levels of clone 709G4 [EGL nine homologue 3 (EGLN3)] in various normal and tumor tissues analyzed with either cDNA microarrays or quantitative real-time PCR. Expression levels of microarray experiments are presented as the ratio of intensities of Cy3 (green fluorescence; individual probe) versus Cy5 (red fluorescence; pool of critical normal tissues); data from real-time PCR are shown as relative copy numbers (reference: β -actin). The different tissue types are distinguished with the indicated colors.

and in Ewing's sarcoma (25). In each case, the use of microarrays to screen the cDNA clones isolated by subtraction cloning resulted in an exceptionally high fraction of differentially expressed cDNA clones present on the arrays and allowed the identification of previously unknown genes. In each of these studies, a clone collection of several hundred cDNAs has been generated from an individual type of tumor or cell line, as appropriate to each question addressed. As a major advance, in the present study, we have assembled a large collection of ~ 9250 clones from seven subtractive cDNA libraries derived from four different tissues of origin (Table 1). As the majority of cDNA clones were exclusively derived by subtractive gene cloning of tumor *versus* normal, the library has been significantly enriched for tumor-specific gene expression. Because of its large size and diverse and balanced sources, it seems likely that most genes expressed in the tumor entities analyzed here are represented in this cDNA clone collection. To our knowledge, this clone collection is the largest one derived by subtraction cloning published to date and one of the

largest human tumor-specific cDNA libraries available. For the production of cDNA microarrays, we have further added 1682 individually selected I.M.A.G.E. clones with a known or suspected role in tumor formation, as well as 105 individually cloned genes previously found to be at least 6-fold up-regulated in colon cancer *versus* normal colon (N. Schweifer, personal communication).

For the preparation of our subtractive libraries, we have developed and applied a protocol for the generation of cDNA fragments with increased size. When following the original protocol relying on a restriction enzyme recognizing only a 4-base motif such as *RsaI* (14), we have observed a high percentage of fragments to be smaller than 50 bp. Accordingly, we have used a set of 6-base recognizing restriction enzymes, 3 with A/T-rich (primarily found at the 3'-end of eukaryotic cDNAs) and 3 with G/C-rich recognition sequences (characteristic for 5'-termini of eukaryotic genes). As shown in Fig. 1, our approach resulted in a considerable shift toward longer cDNA fragments; sequence analysis of several thousands of clones revealed an increase in average length to ~ 800 bp. Such longer cDNA fragments

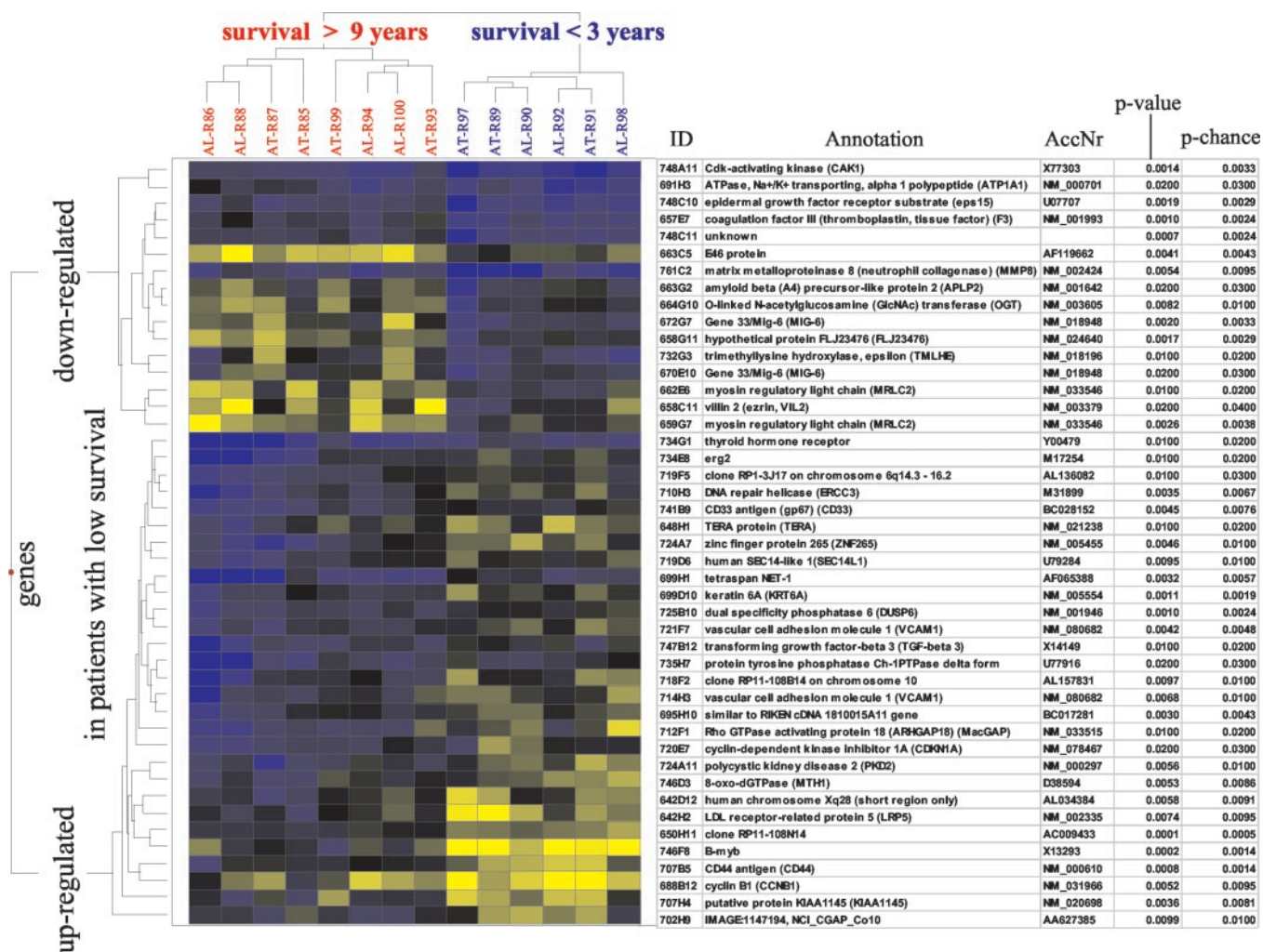


Fig. 6. Gene expression correlates with long or short overall survival of breast cancer patients. The table lists 42 genes with an expression profile most highly associated with the two survival groups (>9 versus <3 years), as selected by Student's t test (significance, $P < 0.02$), accompanied by P-chance analysis, as described in "Materials and Methods," to eliminate false positives. Only genes with $P < P$ -chance were selected. Expression values of the listed genes, up-regulated or down-regulated in low or high survivors, respectively, are visualized as heatmap obtained by hierarchical clustering of genes and samples.

are more favorable for cDNA microarrays (26): they warrant efficient hybridization; minimize cross-hybridization (a problem often observed with short probes); and facilitate annotation after sequencing. Furthermore, our cDNA fragments typically are of sufficient length to be directly used in follow-up studies, *e.g.*, involving translation of portions of the respective gene, circumventing the tedious recloning of longer cDNA fragments of the genes of interest.

Quality Control of RNA Samples and cDNA Microarrays. Intermingled nontumor cells, particularly residual normal cells and infiltrating leukocytes, may heavily interfere with tumor-specific transcription profiles. Therefore, tissue specimen were characterized before RNA isolation. Cryosections were taken to ensure integrity of tumor tissue, to confirm histological typing of tumor samples, and to assess the percentage of tumor cells, residual normal cells, hyperplastic cells, necrotic areas, and infiltrating leukocytes (Table 2). On the basis of these criteria, only 50% of tumor samples were used for RNA isolation. Next, RNA preparations were subjected to capillary electrophoresis (Bioanalyzer) for quality control and quantification, being more accurate than standard photometric assays, particularly for low amounts of mRNA. On the basis of these studies together with data derived from PCR analyses, an additional 30% of RNA preparations was excluded from further analysis due to inadequate RNA quality or due to the presence of chromosomal DNA contamination, as deter-

mined by PCR analysis. We consider this stringent selection procedure as essential because it directly influences hybridization efficiency, reproducibility, and statistical analyses of the results.

To confirm the results of our microarray experiments with an independent method, 5 randomly chosen genes (N-myc downstream regulated, *OSF-2*, *TP73L*, *EGLN3*, and *NAT1*; Table 3) were subjected to real-time RT-PCR. The expression levels quantified by real-time RT-PCR highly correlated with those determined by our cDNA chip analysis, confirming the reliability of our microarray experiments (Fig. 5). In these analyses, we used β -actin as a reference gene because it was the gene with the least fluctuation in our samples according to the chip data. In contrast, *GAPDH* (frequently used for normalization) was found to be much higher expressed in the liver than in any other tissue.

Tissue-Wide Expression Profiling. In contrast to most, if not all, previously reported gene expression profiling studies of human cancer, we compared the expression profiles of representative tumor samples to those of 16 different critical normal tissues, including all normal tissues corresponding to the tumor samples analyzed (tissue-wide expression profile). Accordingly, we used a mRNA-pool of 16 critical normal tissues as a reference probe. In addition, we performed hybridization experiments not only with tumor samples but also

individually with each of the 16 critical normal tissues of the reference pool plus 6 noncritical normal tissues.

Application of restrictive criteria for gene selection resulted in characteristic differences of transcription profiles to those reported previously for solid tumors of the breast, lung, and kidney. For example, because of the extensive comparison to multiple normal tissues, we could eliminate expression signatures derived from infiltrating cells of the immune system such as immunoglobulin genes (B cells), T-cell receptor, *CD3D* (T cells), or lysozyme and chitinase 1 (macrophages/monocytes) from our list of tumor-specific genes (Table 3). All those genes have been included as up-regulated in tumors in previous transcriptional profiling studies, *e.g.*, for breast cancer (5).

Importantly, this tissue-wide expression profiling approach allows us to identify those genes up-regulated in specific tumors with no or low expression in all 16 critical normal tissues tested, which is important for the development of a chemotherapy with less severe or no side effects and which is an absolute prerequisite for immune therapy. The latter approach aims at the induction of a systemic immune response against a tumor-specific/associated antigen (TAA) and—if the TAA is not exclusively expressed in tumors but also in critical normal tissues—may also lead to a destructive autoimmune response generating severe side effects (27, 28). For example, we found prostaglandin D synthetase to be slightly up-regulated in breast cancer but most prominently expressed in several vital tissues such as heart (8.6-fold higher than the reference pool) and brain (6.2-fold higher). The nonreceptor tyrosine kinase *Etk/Bmx*, which was found to be up-regulated ~3-fold in some of the LAC and LSCC samples, has been reported to play an important role for prostate cancer progression and has been suggested as a novel target for chemotherapy in prostate cancer (29). However, our tissue wide expression profiling revealed a strong expression in heart, kidney, and skeletal muscle (6.1-, 5-, and 3-fold higher than in the reference pool, respectively), indicating that severe side effects would have to be expected upon use of *Etk/Bmx* as a novel therapeutic intervention site. Another example is apolipoprotein D, which has been correlated with malignant transformation and poor prognosis in prostate cancer patients (30, 31). Our study revealed overexpression of apolipoprotein D only in breast cancer samples derived from patients with >9 year overall survival. Again, we found strong expression in several critical normal tissues such as brain, heart, and trachea (6.8-, 4.9-, and 3.5-fold higher than the reference probe).

Identification of Tumor Type-Specific Genes. We also identified genes overexpressed in ~100% of a specific cancer type such as vascular endothelial growth factor or insulin-like growth factor binding protein 3 in RCC. Other genes were found to be up-regulated only in a subset of a given tumor type such as stromelysin 3 or thrombospondin 2 in breast cancer. All those selected genes exhibit at least a 2-fold up-regulation in at least 20% of samples of any tumor type compared with the highest expression value in any of the 16 critical normal tissues, which makes them promising putative targets in an anticancer therapy (Table 3). Among the selected candidate genes, we identified a number of genes that have been previously described as tumor markers, *e.g.*, pronapsin A, a gene specifically up-regulated in LAC but absent in LSCC (32) or *NAT-1*, which is involved in detoxification and used as a potential breast cancer marker (33).

Carcinogenesis and Genes Involved in Ca²⁺ Homeostasis and Bone Matrix Mineralization. Serial analysis of gene expression analysis of differentially expressed genes in non-small cell lung cancer (34) and a comparable study of LSCC (24) identified genes that overlap with many genes found in our study such as (*a*) tissue-specific genes such as keratin 6 isoforms, other cytokeratins that have

been documented as potential markers for lung cancer, pemphigus vulgaris antigen, and annexin, (*b*) tumor-specific genes such as parathyroid hormone-related peptide (*PTHrP/PTHLH*), which causes humoral hypercalcaemia associated with malignant tumors such as leukemia, RCC, prostate, and breast cancer, and LSCC (35). Importantly, we have identified several differentially regulated genes that are known to be involved in Ca²⁺ homeostasis; for example, *RCN1*, *CALCA*, and S100 proteins such as S100A10 and S100A11, a subgroup of the EF-hand Ca²⁺-binding protein family (Table 3). In metastatic cell lines, an altered intracellular localization has been demonstrated, supporting the hypothesis that S100 proteins might play a crucial role in the regulation of Ca²⁺ homeostasis in cancer cells (36). For example, S100A2 was found to be highly up-regulated in ovarian cancer together with other members of the S100 protein family (37), whereas an increase of S100A6 expression correlates with an increased malignancy in colon tumors (38). Although a more detailed analysis on the expression profile of members of the S100 protein family in tumors of different origin is needed, this protein class provides promising intervention sites for novel therapeutic strategies.

Together with genes involved in Ca²⁺ homeostasis, we identified genes in breast but also in lung cancer and RCC involved in bone matrix mineralization such as osteonectin (SPARC), osteopontin (SPP1), and OSF-2 (Table 3). Although we have not included osteoclasts and osteoblasts in our panel of normal tissues, the observation that specific tumor types not originating from bone express higher levels of these genes than 16 critical normal tissues is intriguing. Noteworthy, the skeleton is the preferred target of metastatic human breast cancer cells. Bone metastases are indeed found in virtually all advanced breast cancer patients. The high osteotropism of breast cancer cells suggests that they exhibit a selective affinity for mineralized tissues. Mammary malignant cells are able to induce hydroxyapatite crystal deposition within the primary tumor supporting the hypothesis that they can generate a microenvironment that favors the crystallization of calcium and phosphate ions into the bone-specific hydroxyapatite. The ectopic expression of bone matrix proteins in breast cancer could be involved in conferring osteotropic properties to circulating metastatic breast cancer cells and opens the possibility for therapeutic interference with microcalcification during the homing process of metastatic breast cancer cells (39). Interestingly, the osteoclast differentiation/activation factor osteoprotegerin-ligand has been shown to be essential for normal mammary gland development and to be responsible for calcium release from the skeleton required for transmission of maternal calcium to neonates in mammals (40). Therefore, normal cells of the mammary gland may already exhibit some properties of bone remodeling cells, a function that might be recruited/activated in breast tumor cells as well.

Gene Expression Profiles Predicting the Overall Survival of Breast Cancer Patients. We have used tumors of lymph node-positive breast cancer patients with known clinical outcome to determine gene expression signatures predictive of long or short overall survival. Although we have analyzed only a small number of patients, our results led to the identification of novel potential diagnostic marker genes. Furthermore, when taken together with other array studies, our findings highlight the consistent associations of gene expression profiles with clinical outcome (41–44). Several genes found to be overexpressed in patients with short survival (Fig. 6) have already been discussed in the context of breast cancer such as *TGF-β3*, *VCAM-1*, *CD44*, thyroid hormone receptor, and cyclin B1 (45–50), whereas others have not such as *ERG2*, *B-Myb*, *MTH1*, and *NET-1*. Genes down-regulated relative to normal tissues in patients with short survival are, for example, *MIG-6*, *Eps15*, and *APLP2*.

Interestingly, both *MIG-6* and *Eps15* are negative regulators of signaling via the epidermal growth factor receptor, a positive key regulator of breast tumorigenesis (51, 52).

Recently, van 't Veer *et al.* (41) reported on a set of 70 genes with an expression pattern by which breast cancer patients could be classified into those with a poor prognosis and those with a good prognosis with high accuracy. Although these 70 prognostic genes are largely nonidentical to the 42 genes identified by us, many of them are functionally closely related and are involved, *e.g.*, in cell cycle regulation, invasion and metastasis, angiogenesis, and signal transduction. Moreover, despite considerable differences in patient populations and technology platforms used, the study reported here independently arrives at the same general conclusions as van 't Veer *et al.* (41) and van de Vijver *et al.* (42). Apparently, the ability to metastasize to distant sites, which eventually determines the overall survival, is acquired relatively early during multistep tumorigenesis and thus can be diagnosed in the primary tumor several years before these metastases become manifest (42, 43). This ability to form hematogenous (distant) metastases appears to be largely independent of the presence or absence of lymph node metastases. All patients analyzed in the present study had lymph node metastases at the time of diagnosis; nevertheless, 4 of them remained free of distant metastases and disease relapse for at least 9 years of follow-up.

An important clinical question concerns whether prognosis profiling is equally useful for all patients with breast cancer or whether it is limited to the specific subgroup(s) of patients for whom it has already been demonstrated (44). The fact that despite considerable differences in study design, the results reported here support the major conclusions of van 't Veer *et al.* (41) and van de Vijver *et al.* (42) provides supporting evidence that an accurate prediction of clinical outcome based on gene expression profiling could be generally applicable to all breast cancer patients.

In summary, a modified PCR-based cDNA subtraction method allowed the establishment of seven SSH cDNA libraries that subsequently were used for the preparation of cDNA microarrays. Together with 50 samples derived from lung, breast, or renal cell cancer tissues, a panel of 22 samples from normal tissues was hybridized. This detailed tissue-wide expression profiling led to the identification of 130 individual tumor-specific transcripts (527 clones) showing no or very low expression in 16 vital normal tissues. Gene-wise hierarchical clustering of these 130 genes clearly separated the different tumor types. The majority of the identified genes have not yet been brought into context with tumorigenesis such as genes involved in bone matrix mineralization or genes controlling calcium homeostasis (*RCNI*, *CALCA*, and *S100* protein family). Forty-two genes were identified that significantly correlated with the overall survival of breast cancer patients, genes up-regulated in tumors of patients with a poor prognosis such as cyclin B1, *TGF- β 3*, *B-Myb*, and *Erg2*, and genes down-regulated such as *MIG-6*, *Esp15*, and *CAK*.

ACKNOWLEDGMENTS

We thank Frederike Fessel, Atishendu Gupta, Herwig Machat, Roland Varecka, Chung-I Wu, and Sandra Zinke for their excellent technical assistance. We also thank Christoph Klade for provision of the renal cell cancer subtractive cDNA library, Norbert Schweifer for information on colon cancer-specific transcripts, Thomas Woelfel for pancreatic carcinoma cell line MZ.PC2, and Zacharias Rattler for invaluable stimulating discussions.

REFERENCES

- Sporn, M. B. The war on cancer. *Lancet*, 347: 1377–1381, 1996.
- Fedi, P., Tronick, S. R., and Aaronson, S. A. Growth Factors. In: J. F. Holland (ed.), *Cancer Medicine*, pp. 41–64. Baltimore: Williams and Wilkins, 1997.
- Hanahan, D., and Weinberg, R. A. The hallmarks of cancer. *Cell*, 100: 57–70, 2000.

- Kaelin, W. G., Jr. Choosing anticancer drug targets in the postgenomic era. *J. Clin. Invest.*, 104: 1503–1506, 1999.
- Perou, C. M., Sorlie, T., Eisen, M. B., van de, R. M., Jeffrey, S. S., Rees, C. A., Pollack, J. R., Ross, D. T., Johnsen, H., Akslen, L. A., Fluge, O., Pergamenschikov, A., Williams, C., Zhu, S. X., Lonning, P. E., Borresen-Dale, A. L., Brown, P. O., and Botstein, D. Molecular portraits of human breast tumours. *Nature (Lond.)*, 406: 747–752, 2000.
- Takahashi, M., Rhodes, D. R., Furge, K. A., Kanayama, H., Kagawa, S., Haab, B. B., and Teh, B. T. Gene expression profiling of clear cell renal cell carcinoma: gene identification and prognostic classification. *Proc. Natl. Acad. Sci. USA*, 98: 9754–9759, 2001.
- Golub, T. R., Slonim, D. K., Tamayo, P., Huard, C., Gaasenbeek, M., Mesirov, J. P., Coller, H., Loh, M. L., Downing, J. R., Caligiuri, M. A., Bloomfield, C. D., and Lander, E. S. Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. *Science (Wash. DC)*, 286: 531–537, 1999.
- Alizadeh, A. A., Eisen, M. B., Davis, R. E., Ma, C., Lossos, I. S., Rosenwald, A., Boldrick, J. C., Sabet, H., Tran, T., Yu, X., Powell, J. I., Yang, L., Marti, G. E., Moore, T., Hudson, J., Jr., Lu, L., Lewis, D. B., Tibshirani, R., Sherlock, G., Chan, W. C., Greiner, T. C., Weisenburger, D. D., Armitage, J. O., Warnke, R., Levy, R., Wilson, W., Grever, M. R., Byrd, J. C., Botstein, D., Brown, P. O., and Staudt, L. M. Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. *Nature (Lond.)*, 403: 503–511, 2000.
- Bittner, M., Meltzer, P., Chen, Y., Jiang, Y., Seftor, E., Hendrix, M., Radmacher, M., Simon, R., Yakhini, Z., Ben Dor, A., Sampedro, N., Dougherty, E., Wang, E., Marincola, F., Gooden, C., Lueders, J., Glatfelter, A., Pollock, P., Carpten, J., Gillanders, E., Leja, D., Dietrich, K., Beaudry, C., Berens, M., Alberts, D., and Sondak, V. Molecular classification of cutaneous malignant melanoma by gene expression profiling. *Nature (Lond.)*, 406: 536–540, 2000.
- Hedenfalk, I., Duggan, D., Chen, Y., Radmacher, M., Bittner, M., Simon, R., Meltzer, P., Gusterson, B., Esteller, M., Kallioniemi, O. P., Wilfond, B., Borg, A., and Trent, J. Gene-expression profiles in hereditary breast cancer. *N. Engl. J. Med.*, 344: 539–548, 2001.
- Bangur, C. S., Switzer, A., Fan, L., Marton, M. J., Meyer, M. R., and Wang, T. Identification of genes over-expressed in small cell lung carcinoma using suppression subtractive hybridization and cDNA microarray expression analysis. *Oncogene*, 21: 3814–3825, 2002.
- Bouras, T., Southey, M. C., Chang, A. C., Reddel, R. R., Willhite, D., Glynne, R., Henderson, M. A., Armes, J. E., and Venter, D. J. Stanniocalcin 2 is an estrogen-responsive gene coexpressed with the estrogen receptor in human breast cancer. *Cancer Res.*, 62: 1289–1295, 2002.
- Buckley, C. D., Pilling, D., Henriquez, N. V., Parsonage, G., Threlfall, K., Scheel-Toellner, D., Simmons, D. L., Akbar, A. N., Lord, J. M., and Salmon, M. RGD peptides induce apoptosis by direct caspase-3 activation. *Nature (Lond.)*, 397: 534–539, 1999.
- Diatchenko, L., Lau, Y. F., Campbell, A. P., Chenchik, A., Moqadam, F., Huang, B., Lukyanov, S., Lukyanov, K., Gurskaya, N., Sverdlov, E. D., and Siebert, P. D. Suppression subtractive hybridization: a method for generating differentially regulated or tissue-specific cDNA probes and libraries. *Proc. Natl. Acad. Sci. USA*, 93: 6025–6030, 1996.
- Sambrook, J., and Russell, D. W. *Molecular Cloning*, Ed. 3. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press, 2001.
- Eisen, M. B., and Brown, P. O. DNA arrays for analysis of gene expression. *Methods Enzymol.*, 303: 179–205, 1999.
- Schena, M., Shalon, D., Davis, R. W., and Brown, P. O. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science (Wash. DC)*, 270: 467–470, 1995.
- Dudoit, S., Yang, Y. H., Speed, T. P., and Callow, M. J. Statistical methods for identifying differentially expressed genes in replicated cDNA microarray experiments. *Statistica Sinica*, 12: 111–139, 2002.
- Takahashi, K., Paladini, R. D., and Coulombe, P. A. Cloning and characterization of multiple human genes and cDNAs encoding highly related type II keratin 6 isoforms. *J. Biol. Chem.*, 270: 18581–18592, 1995.
- Klade, C. S., Dohnal, A., Furst, W., Sommergruber, W., Heider, K. H., Gharwan, H., Ratschek, M., and Adolf, G. R. Identification and characterization of 9D7, a novel human protein overexpressed in renal cell carcinoma. *Int. J. Cancer*, 97: 217–224, 2002.
- Epstein, A. C., Gleadle, J. M., McNeill, L. A., Hewitson, K. S., O'Rourke, J., Mole, D. R., Mukherji, M., Metzen, E., Wilson, M. I., Dhanda, A., Tian, Y. M., Masson, N., Hamilton, D. L., Jaakkola, P., Barstead, R., Hodgkin, J., Maxwell, P. H., Pugh, C. W., Schofield, C. J., and Ratcliffe, P. J. *C. elegans* EGL-9 and mammalian homologs define a family of dioxygenases that regulate HIF by prolyl hydroxylation. *Cell*, 107: 43–54, 2001.
- Yang, G. P., Ross, D. T., Kuang, W. W., Brown, P. O., and Weigel, R. J. Combining SSH and cDNA microarrays for rapid identification of differentially expressed genes. *Nucleic Acids Res.*, 27: 1517–1523, 1999.
- Xu, J., Stolk, J. A., Zhang, X., Silva, S. J., Houghton, R. L., Matsumura, M., Vedvick, T. S., Leslie, K. B., Badaro, R., and Reed, S. G. Identification of differentially expressed genes in human prostate cancer using subtraction and microarray. *Cancer Res.*, 60: 1677–1682, 2000.
- Wang, T., Hopkins, D., Schmidt, C., Silva, S., Houghton, R., Takita, H., Repasky, E., and Reed, S. G. Identification of genes differentially over-expressed in lung squamous cell carcinoma using combination of cDNA subtraction and microarray analysis. *Oncogene*, 19: 1519–1528, 2000.
- Welford, S. M., Gregg, J., Chen, E., Garrison, D., Sorensen, P. H., Denny, C. T., and Nelson, S. F. Detection of differentially expressed genes in primary tumor tissues

- using representational differences analysis coupled to microarray hybridization. *Nucleic Acids Res.*, *26*: 3059–3065, 1998.
26. Stillman, B. A., and Tonkinson, J. L. Expression microarray hybridization kinetics depend on length of the immobilized DNA but are independent of immobilization substrate. *Anal. Biochem.*, *295*: 149–157, 2001.
 27. Streilein, J. W. Unraveling immune privilege. *Science (Wash. DC)*, *270*: 1158–1159, 1995.
 28. Pardoll, D. M. Inducing autoimmune disease to treat cancer. *Proc. Natl. Acad. Sci. USA*, *96*: 5340–5342, 1999.
 29. Lee, L. F., Guan, J., Qiu, Y., and Kung, H. J. Neuropeptide-induced androgen independence in prostate cancer cells: roles of nonreceptor tyrosine kinases Etk/Bmx, Src, and focal adhesion kinase. *Mol. Cell Biol.*, *21*: 8385–8397, 2001.
 30. Lopez-Boado, Y. S., Puente, X. S., Alvarez, S., Tolivia, J., Binderup, L., and Lopez-Otin, C. Growth inhibition of human breast cancer cells by 1,25-dihydroxyvitamin D₃ is accompanied by induction of apolipoprotein D expression. *Cancer Res.*, *57*: 4091–4097, 1997.
 31. Zhang, S. X., Bentel, J. M., Ricciardelli, C., Horsfall, D. J., Haagensen, D. E., Marshall, V. R., and Tilley, W. D. Immunolocalization of apolipoprotein D, androgen receptor and prostate specific antigen in early stage prostate cancers. *J. Urol.*, *159*: 548–554, 1998.
 32. Chuman, Y., Bergman, A., Ueno, T., Saito, S., Sakaguchi, K., Alaiya, A. A., Franzen, B., Bergman, T., Arnott, D., Auer, G., Appella, E., Jornvall, H., and Linder, S. Napsin A, a member of the aspartic protease family, is abundantly expressed in normal lung and kidney tissue and is expressed in lung adenocarcinomas. *FEBS Lett.*, *462*: 129–134, 1999.
 33. Krajcinovic, M., Ghadirian, P., Richer, C., Sinnett, H., Gandini, S., Perret, C., Lacroix, A., Labuda, D., and Sinnett, D. Genetic susceptibility to breast cancer in French-Canadians: role of carcinogen-metabolizing enzymes and gene-environment interactions. *Int. J. Cancer*, *92*: 220–225, 2001.
 34. Hibi, K., Liu, Q., Beaudry, G. A., Madden, S. L., Westra, W. H., Wehage, S. L., Yang, S. C., Heitmiller, R. F., Bertelsen, A. H., Sidransky, D., and Jen, J. Serial analysis of gene expression in non-small cell lung cancer. *Cancer Res.*, *58*: 5690–5694, 1998.
 35. Davidson, L. A., Black, M., Carey, F. A., Logue, F., and McNicol, A. M. Lung tumours immunoreactive for parathyroid hormone related peptide: analysis of serum calcium levels and tumour type. *J. Pathol.*, *178*: 398–401, 1996.
 36. Mueller, A., Bachi, T., Hochli, M., Schafer, B. W., and Heizmann, C. W. Subcellular distribution of S100 proteins in tumor cells and their relocation in response to calcium activation. *Histochem. Cell Biol.*, *111*: 453–459, 1999.
 37. Hough, C. D., Cho, K. R., Zonderman, A. B., Schwartz, D. R., and Morin, P. J. Coordinately up-regulated genes in ovarian cancer. *Cancer Res.*, *61*: 3869–3876, 2001.
 38. Bronckart, Y., Decaestecker, C., Nagy, N., Harper, L., Schafer, B. W., Salmon, I., Pochet, R., Kiss, R., and Heizman, C. W. Development and progression of malignancy in human colon tissues are correlated with expression of specific Ca(2+)-binding S100 proteins. *Histol. Histopathol.*, *16*: 707–712, 2001.
 39. Bellahcene, A., and Castronovo, V. Expression of bone matrix proteins in human breast cancer: potential roles in microcalcification formation and in the genesis of bone metastases. *Bull. Cancer*, *84*: 17–24, 1997.
 40. Fata, J. M., Kong, Y.-Y., Li, J., Sassaki, T., Irie-Sasaki, J., Moorehead, R. A., Elliott, R., Scully, S., Voura, E. B., Lacey, D. L., Boyle, W. J., Khokha, R., and Penninger, J. M. The osteoclast differentiation factor osteoprotegerin-ligand is essential for mammary gland development. *Cell*, *103*: 41–50, 2000.
 41. van 't Veer, L. J., Dai, H., van de Vijver, M. J., He, Y. D., Hart, A. A., Mao, M., Peterse, H. L., van der, K. K., Marton, M. J., Witteveen, A. T., Schreiber, G. J., Kerkhoven, R. M., Roberts, C., Linsley, P. S., Bernards, R., and Friend, S. H. Gene expression profiling predicts clinical outcome of breast cancer. *Nature (Lond.)*, *415*: 530–536, 2002.
 42. van de Vijver, M. J., He, Y. D., van't Veer, L. J., Dai, H., Hart, A. A., Voskuil, D. W., Schreiber, G. J., Peterse, J. L., Roberts, C., Marton, M. J., Parrish, M., Atsma, D., Witteveen, A., Glas, A., Delahaye, L., van der Velde, T., Bartelink, H., Rodenhuis, S., Rutgers, E. T., Friend, S. H., and Bernards, R. A gene-expression signature as a predictor of survival in breast cancer. *N. Engl. J. Med.*, *347*: 1999–2009, 2002.
 43. Bernards, R., and Weinberg, R. A. A progression puzzle. *Nature (Lond.)*, *418*: 823, 2002.
 44. Kallioniemi, A. Molecular signatures of breast cancer: predicting the future. *N. Engl. J. Med.*, *347*: 2067–2068, 2002.
 45. Ghellal, A., Li, C., Hayes, M., Byrne, G., Bundred, N., and Kumar, S. Prognostic significance of TGF β 1 and TGF β 3 in human breast carcinoma. *Anticancer Res.*, *20*: 4413–4418, 2000.
 46. Byrne, G. J., and Bundred, N. J. Surrogate markers of tumoral angiogenesis. *Int. J. Biol. Markers*, *15*: 334–339, 2000.
 47. Herrera-Gayol, A., and Jothy, S. Adhesion proteins in the biology of breast cancer: contribution of CD44. *Exp. Mol. Pathol.*, *66*: 149–156, 1999.
 48. Shao, Z. M., Sheikh, M. S., Rishi, A. K., Dawson, M. I., Li, X. S., Wilber, J. F., Feng, P., and Fontana, J. A. Thyroid hormone enhancement of estradiol stimulation of breast carcinoma proliferation. *Exp. Cell Res.*, *218*: 1–8, 1995.
 49. Barrett, K. L., Demiranda, D., and Katula, K. S. Cyclin b1 promoter activity and functional cdk1 complex formation in G₁ phase of human breast cancer cells. *Cell Biol. Int.*, *26*: 19–28, 2002.
 50. Megha, T., Lazzi, S., Ferrari, F., Vatti, R., Howard, C. M., Cevenini, G., Leoncini, L., Luzi, P., Giordano, A., and Tosi, P. Expression of the G₂-M checkpoint regulators cyclin B1 and P34CDC2 in breast cancer: a correlation with cellular kinetics. *Anticancer Res.*, *19*: 163–169, 1999.
 51. Tang, C. K., Gong, X. Q., Moscatello, D. K., Wong, A. J., and Lippman, M. E. Epidermal growth factor receptor vIII enhances tumorigenicity in human breast cancer. *Cancer Res.*, *60*: 3081–3087, 2000.
 52. Kim, H., and Muller, W. J. The role of the epidermal growth factor receptor family in mammary tumorigenesis and metastasis. *Exp. Cell Res.*, *253*: 78–87, 1999. R. E., Nairn, A. C., Greengard, P., Gandy, S. E., and Kirino, Y. Phosphorylation of Alzheimer β -amyloid precursor-like proteins. *Biochemistry*, *36*: 4643–4649, 1997.
 53. Sobin, L. H., and Wittekind, W. Ch. International Union Against Cancer. TNM Classification of Malignant Tumors, Ed. 5th. New York: John Wiley & Sons, 1997.

Cancer Research

The Journal of Cancer Research (1916–1930) | The American Journal of Cancer (1931–1940)

Tissue-Wide Expression Profiling Using cDNA Subtraction and Microarrays to Identify Tumor-Specific Genes

Stefan Amatschek, Ulrich Koenig, Herbert Auer, et al.

Cancer Res 2004;64:844-856.

Updated version Access the most recent version of this article at:
<http://cancerres.aacrjournals.org/content/64/3/844>

Supplementary Material Access the most recent supplemental material at:
<http://cancerres.aacrjournals.org/content/suppl/2004/11/01/64.3.844.DC1>

Cited articles This article cites 49 articles, 16 of which you can access for free at:
<http://cancerres.aacrjournals.org/content/64/3/844.full.html#ref-list-1>

Citing articles This article has been cited by 30 HighWire-hosted articles. Access the articles at:
</content/64/3/844.full.html#related-urls>

E-mail alerts [Sign up to receive free email-alerts](#) related to this article or journal.

Reprints and Subscriptions To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.

Permissions To request permission to re-use all or part of this article, contact the AACR Publications Department at permissions@aacr.org.